

#4

## BEST AVAILABLE COPY

OIPE

## RAW SEQUENCE LISTING

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:03

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

## SEQUENCE LISTING

ENTERED

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: MAERTENS, GEERT

6 BOSMAN, FONS

7 DE MARTYNOFF, GUY

8 BUYSE, MARIE-ANGE

10 (ii) TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

11 PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

13 (iii) NUMBER OF SEQUENCES: 111

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: NIXON &amp; VANDERHYE P.C.

17 (B) STREET: 1100 NORTH GLEBE ROAD

18 (C) CITY: ARLINGTON

19 (D) STATE: VIRGINIA

20 (E) COUNTRY: U.S.A.

21 (F) ZIP: 22201-4714

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Floppy disk

25 (B) COMPUTER: IBM PC compatible

26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

29 (vi) CURRENT APPLICATION DATA:

C--&gt; 30 (A) APPLICATION NUMBER: US/09/899,303A

C--&gt; 31 (B) FILING DATE: 06-Jul-2001

32 (C) CLASSIFICATION:

34 (vii) PRIOR APPLICATION DATA:

35 (A) APPLICATION NUMBER: US/08/612,973

36 (B) FILING DATE: 11-MAR-1996

38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: BYRNE, THOMAS E.

40 (B) REGISTRATION NUMBER: 32,205

41 (C) REFERENCE/DOCKET NUMBER: 1487-10

43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: (703) 816-4000

45 (B) TELEFAX: (703) 816-4100

49 (2) INFORMATION FOR SEQ ID NO: 1:

51 (i) SEQUENCE CHARACTERISTICS:

52 (A) LENGTH: 21 base pairs

53 (B) TYPE: nucleic acid

54 (C) STRANDEDNESS: single

55 (D) TOPOLOGY: linear

57 (ii) MOLECULE TYPE: cDNA

59 (iii) HYPOTHETICAL: NO

C--&gt; 61 (iv) ANTI-SENSE: NO

66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

68 GGCATGCAAG CTTAATTAAT T

70 (2) INFORMATION FOR SEQ ID NO: 2:

21

## RAW SEQUENCE LISTING

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:03

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

```

72      (i) SEQUENCE CHARACTERISTICS:
73          (A) LENGTH: 68 base pairs
74          (B) TYPE: nucleic acid
75          (C) STRANDEDNESS: single
76          (D) TOPOLOGY: linear
78      (ii) MOLECULE TYPE: cDNA
80      (iii) HYPOTHETICAL: NO
C--> 82      (iv) ANTI-SENSE: NO
86      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
88 CCGGGGAGGC CTGCACGTGA TCGAGGGCAG ACACCATCAC CACCATCACT AATAGTTAAT      60
90 TAACTGCA                                          68
92 (2) INFORMATION FOR SEQ ID NO: 3:
94      (i) SEQUENCE CHARACTERISTICS:
95          (A) LENGTH: 642 base pairs
96          (B) TYPE: nucleic acid
97          (C) STRANDEDNESS: single
98          (D) TOPOLOGY: linear
100     (ii) MOLECULE TYPE: cDNA
102     (iii) HYPOTHETICAL: NO
C--> 104     (iv) ANTI-SENSE: NO
107     (ix) FEATURE:
108         (A) NAME/KEY: CDS
109         (B) LOCATION: 1..639
111     (ix) FEATURE:
112         (A) NAME/KEY: mat_peptide
113         (B) LOCATION: 1..636
116     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
118 ATG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTA CTG TCC TGT      48
119 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
120 1          5          10          15
122 CTG ACC ATT CCA GCT TCC GCT TAT GAG GTG CGC AAC GTG TCC GGG ATG      96
123 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
124          20          25          30
126 TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA      144
127 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
128          35          40          45
130 GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT CGG GAG      192
131 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
132          50          55          60
134 AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG CTC GCA GCT      240
135 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
136 65          70          75          80
138 AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC GTC GAT TTG      288
139 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
140          85          90          95
142 CTC GTT GGG GCG GCT GCT CTC TGT TCC GCT ATG TAC GTG GGG GAT CTC      336
143 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
144          100          105          110
146 TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC TCG CCT CGC      384

```

## RAW SEQUENCE LISTING

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:03

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

```

147 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
148      115      120      125
150 CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT CCC GGC CAC      432
151 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
152      130      135      140
154 ATA ACA GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC TGG TCG CCT      480
155 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
156 145      150      155      160
158 ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA CAA GCT GTC      528
159 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
160      165      170      175
162 GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG GGC CTC GCC      576
163 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
164      180      185      190
166 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA      624
167 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
168      195      200      205
170 CTC TTT GCT CTC TAATAG      642
171 Leu Phe Ala Leu
172      210
175 (2) INFORMATION FOR SEQ ID NO: 4:
177 (i) SEQUENCE CHARACTERISTICS:
178 (A) LENGTH: 212 amino acids
179 (B) TYPE: amino acid
180 (D) TOPOLOGY: linear
182 (ii) MOLECULE TYPE: protein
184 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
186 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
187 1      5      10      15
189 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
190      20      25      30
192 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
193      35      40      45
195 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
196      50      55      60
198 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
199 65      70      75      80
201 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
202      85      90      95
204 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
205      100      105      110
207 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
208      115      120      125
210 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
211      130      135      140
213 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
214 145      150      155      160
216 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
217      165      170      175

```

## RAW SEQUENCE LISTING

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:03

Input Set : N:\Crif3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

219 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala  
 220 180 185 190  
 222 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu  
 223 195 200 205  
 225 Leu Phe Ala Leu  
 226 210

229 (2) INFORMATION FOR SEQ ID NO: 5:

231 (i) SEQUENCE CHARACTERISTICS:

232 (A) LENGTH: 795 base pairs

233 (B) TYPE: nucleic acid

234 (C) STRANDEDNESS: single

235 (D) TOPOLOGY: linear

237 (ii) MOLECULE TYPE: cDNA

239 (iii) HYPOTHETICAL: NO

C--&gt; 241 (iv) ANTI-SENSE: NO

244 (ix) FEATURE:

245 (A) NAME/KEY: CDS

246 (B) LOCATION: 1..792

248 (ix) FEATURE:

249 (A) NAME/KEY: mat\_peptide

250 (B) LOCATION: 1..789

253 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

255	ATG TTG GGT AAG GTC ATC GAT ACC CTT ACA TGC GGC TTC GCC GAC CTC	48
256	Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu	
257	1 5 10 15	
259	GTG GGG TAC ATT CCG CTC GTC GGC GCC CCC CTA GGG GGC GCT GCC AGG	96
260	Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg	
261	20 25 30	
263	GCC CTG GCG CAT GGC GTC CGG GTT CTG GAG GAC GGC GTG AAC TAT GCA	144
264	Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala	
265	35 40 45	
267	ACA GGG AAT TTG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG	192
268	Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu	
269	50 55 60	
271	CTG TCC TGT CTG ACC GTT CCA GCT TCC GCT TAT GAA GTG CGC AAC GTG	240
272	Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val	
273	65 70 75 80	
275	TCC GGG ATG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG	288
276	Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val	
277	85 90 95	
279	TAT GAG GCA GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC	336
280	Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys	
281	100 105 110	
283	GTT CGG GAG AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG	384
284	Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr	
285	115 120 125	
287	CTC GCA GCT AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC	432
288	Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His	
289	130 135 140	

## RAW SEQUENCE LISTING

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:03

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

```

291 GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGT TCC GCT ATG TAC GTG      480
292 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
293 145                      150                      155                      160
295 GGG GAC CTC TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC      528
296 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile
297                      165                      170                      175
299 TCG CCT CGC CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT      576
300 Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr
301                      180                      185                      190
303 CCC GGC CAC ATA ACG GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC      624
304 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
305                      195                      200                      205
307 TGG TCG CCT ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA      672
308 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
309                      210                      215                      220
311 CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG      720
312 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
313 225                      230                      235                      240
315 GGT CTC GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT      768
316 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
317                      245                      250                      255
319 GTG ATG CTA CTC TTT GCT CCC TAATAG      795
320 Val Met Leu Leu Phe Ala Pro
321                      260
324 (2) INFORMATION FOR SEQ ID NO: 6:
326 (i) SEQUENCE CHARACTERISTICS:
327 (A) LENGTH: 263 amino acids
328 (B) TYPE: amino acid
329 (D) TOPOLOGY: linear
331 (ii) MOLECULE TYPE: protein
333 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
335 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
336 1                      5                      10                      15
338 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
339                      20                      25                      30
341 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
342                      35                      40                      45
344 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
345                      50                      55                      60
347 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
348 65                      70                      75                      80
350 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
351                      85                      90                      95
353 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
354                      100                     105                     110
356 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr
357                      115                     120                     125
359 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
360                      130                     135                     140

```

## VERIFICATION SUMMARY

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:04

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:61 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:82 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:104 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:241 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:399 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:467 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:536 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:587 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9  
L:644 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:753 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:891 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:911 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:931 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:951 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:971 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:991 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1011 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1083 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:21  
L:1155 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1278 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1408 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1545 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1681 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1818 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1838 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1858 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2113 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2305 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2498 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2676 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2856 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3098 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3444 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3660 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:49  
L:4724 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4744 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4766 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4787 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4807 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4827 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4847 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4867 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4887 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4907 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4927 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4947 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]

## VERIFICATION SUMMARY

DATE: 11/23/2001

PATENT APPLICATION: US/09/899,303A

TIME: 18:16:04

Input Set : N:\Crf3\RULE60\09899303A.raw

Output Set: N:\CRF3\11232001\I899303A.raw

L:4967 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4987 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5007 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5027 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5047 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5067 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]